Position paper for the W3C Do Not Track Workshop

Aleecia M. McDonald

Summary

Preliminary research suggests that user's expectations for Do Not Track (DNT) will not match implementations. While we might imagine changing DNT implementations to align more closely with expectations, it is quite unlikely DNT will change enough to meet user expectations. For example, if users think Do Not Track means no data collection at all, advertisers are unlikely to forgo counting unique clickthrough rates for billing. Furthermore, it is likely there will be multiple approaches to what Do Not Track means in practice, creating additional user confusion and uncertainty. Communicating with users to explain the gap between their expectations and reality is crucial. That means creating mechanisms to support (or at least not preclude) DNT implementers explaining how they implement DNT, and what their implementation means to their users. Unfortunately, current standards proposals do not envision this type of feedback to users. I hope to spark discussion about expanding DNT standards to include the data required to communicate with users.

Established Issues

Users do not understand the Network Advertising Initiative (NAI) description of their members' opt-out cookies. In research from Carnegie Mellon's CUPS laboratory, we presented a screenshot of the NAI website and found only 11% of study respondents selected the correct multiple-choice description of NAI opt-out cookies. Our largest group of respondents mistakenly believed their data would not be collected if they opted out.[1]

Part of the confusion with NAI opt-outs may stem from the multiple ways in which NAI members implement opt-outs. Some OBA companies stop collecting data when they read opt-out cookies. Some companies, including Google, aggregate data from all users who opt-out. Some companies, including Yahoo!, do not change their data collection practices. They stop showing ads tailored based on user data, but data collection continues unchanged. So much variation in outcomes poses a difficult communication problem. There is no one, simple answer to the basic question: what does an opt-out cookie do?

---

[1] McDonald, A. M., and Cranor, L. F. Beliefs and behaviors: Internet users' understanding of behavioral advertising. In *38th Research Conference on Communication, Information and Internet Policy* (Telecommunications Policy Research Conference) (October 2 2010).

In addition, when users see a checkbox labeled "opt out" next an advertiser's name, they are likely to expect they are opting out of seeing advertising from that advertiser. NAI takes great pains to stress that users will see the same number of ads with or without opting out, perhaps because NAI had discovered this is a common misconception. But even with a warning in bold that opt-outs do not reduce ads, we still found that was a common misconception. It is even more difficult to communicate clearly when users hold an expectation that does not match the implementation for privacy controls.

There are three ways in which NAI opt-out cookies research is directly relevant to Do Not Track. First, Google's Chrome browser uses opt-out cookies as their Do Not Track solution. Presumably they have similar communication challenges with their users as the NAI has had. Second, the Do Not Track header sent by both Firefox and Internet Explorer will likely to encompass multiple implementations, as different parties define "tracking" in different ways. Most immediately, some companies may initially treat the DNT header exactly as they do opt-out cookies, thus recreating all of the ambiguity already inherent in opt-out cookies. Third, preliminary research strongly suggests when users see the phrase "Do Not Track," they mistakenly believe this means all data collection stops. As with opt-out cookies, when users think they understand what something means, but it turns out to mean something else, there is a challenge to communicate across the gap between user expectations and reality.

Mind the Gap

As just one example of how complicated defining "tracking" has become, Figure 1 contains a list of data uses that the Center for Democracy and Technology consider to be tracking, or not.[2] To understand this chart, users would need to understand at least the difference between first- and third-party websites, what behavioral advertising is, the types of data collected for behavioral advertising, the difference between identifiable and non-identifiable data, reporting, and analytics.

In a pilot test for a larger on-going research study, I found a majority of users expect Do Not Track to eliminate all data collection. The study starts by asking participants what they expect a Do Not Track button in their web browser would do. Participants work from their own expectations rather than a definition of DNT. They check the types of data they believe can be collected before and after clicking a Do Not Track button, with a subset of results shown in Figure 2.

---

[2] Center for Democracy & Technology. What does "Do Not Track" mean? A scoping proposal by the Center for Democracy & Technology, January 2011. http://cdt.org/files/pdfs/ CDT-DNT-Report.pdf.

Tracking	Not tracking
Third-party online behavioral advertising	Third-party ad and content delivery
Third-party behavioral data collection for first party uses	Third-party reporting
Third-party behavioral data collection for other uses	Third-party analytics
Behavioral data collected by first parties and transferred to third parties in identifiable form	Third-party contextual advertising
	First-party data collection and use
	Federated identity transaction data
	Data collection required by law and for legitimate fraud prevention purposes

Figure 1: The Center for Democracy & Technology's list of examples of data used for tracking and not tracking, illustrating their definition of tracking

To highlight a few of the more interesting results in the pilot study:

- 61% of respondents expected that if they clicked a Do Not Track button, websites would collect no data at all. None of the current proposals for Do Not Track contemplate limiting data collection to nothing for first party use, yet that is what many users expect from Do Not Track.
- Respondents did not expect Do Not Track to work by aggregating their data with other user's data, with only 5% selecting that as a possibility, yet this is how some companies treat opt-out cookies today. Similarly, participants did not expect Do Not Track to work by collecting the same information, but anonymizing it, with only 7% selecting that as a possibility. One reason participants may not expect DNT to protect privacy via aggregation is because they believe that is already how the Internet currently works, and do not understand that they are uniquely identified today.
- Only 7% of respondents expected that websites could collect the same data before and after users click Do Not Track. Some DNT implementations may limit data use rather than data collection, as Yahoo! does with opt-out cookies today.

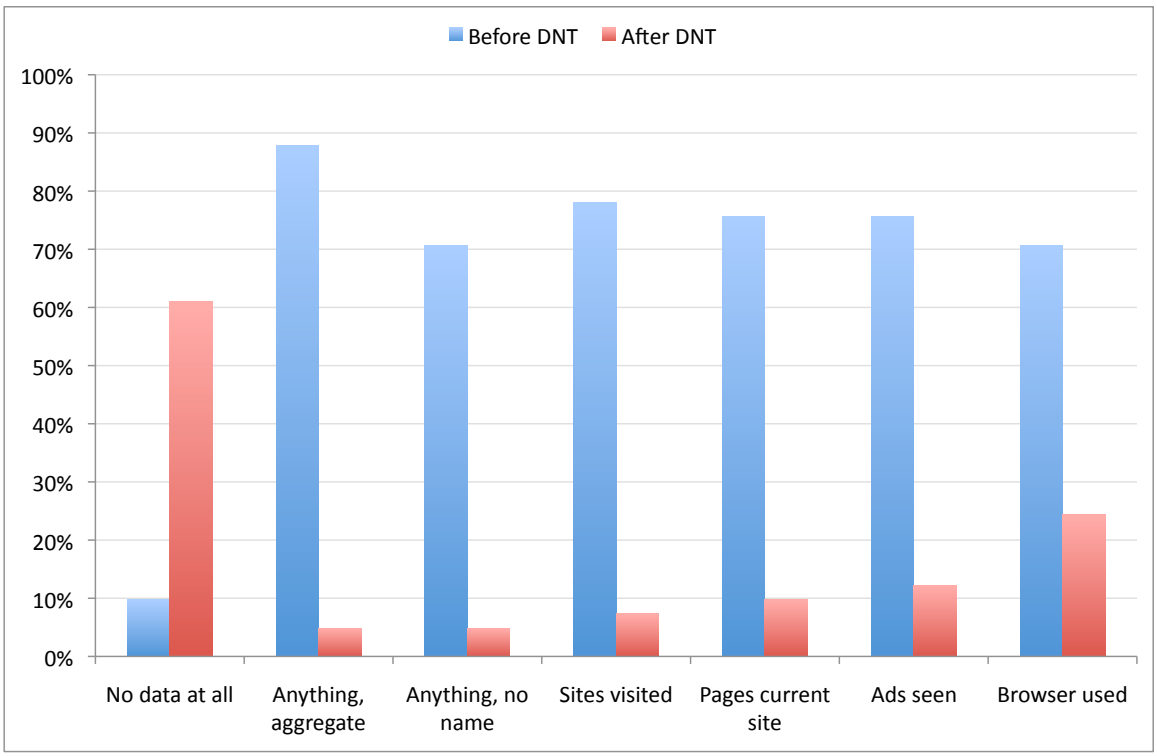

Figure 2: Types of data users think websites can collect, before (in blue) and after (in red) clicking a "Do Not Track" button in their web browser

A larger study is currently underway, and results will be ready for discussion at the W3C workshop. The high level conclusion should remain stable: when users hear "Do Not Track," the majority of users believe data collection stops.

The Role for Standards

One way to address the gap between user expectations and reality is to communicate what DNT actually does. Existing mechanisms include online help files from browser makers and privacy policies from DNT implementers. However, most users do not read online help or privacy policies, and do not understand the number of entities collecting data about them on any given website.

The current DNT proposal before the IETF contemplates browsers sending a DNT header, and receiving confirmation of what the browser sent.[3] Beyond simple acknowledgment of the DNT header, this standard omits any automated mechanism for DNT implementers to communicate with end users. If instead standards build in communication channels, we can create transparency around what DNT means for any given DNT implementer. With multiple DNT implementers rolled up together, we can provide a holistic view of privacy implications for a particular visit to a

[3] J. Mayer, A. Narayanan, S. Stamm. Do Not Track: A Universal Third-Party Web Tracking Opt Out, IETF Draft (March, 2011).

particular website. Finally, we can create an opportunity for sites to communicate the benefits of personalization and why they use data.

If the W3C were able to agree upon standards for communication about DNT implementation details, IETF might extend the current proposal. Or, there may be better forms of communication that do not require modifying HTTP headers. As an example, an extended standard DNT response might include:

- An acknowledgement of the DNT header, as currently proposed to the IETF
- A URL with human-readable text describing what Do Not Track means to that particular entity. This could be as a simple as an anchored tag in a privacy policy, for example www.acme.com/privacy.html#dnt
- A standardized code describing DNT practices, as below

While privacy policies have too much variation to fit neatly into a handful of pre-defined categories, DNT implementations may be more tractable. For example, a site might be classified as type-0 if it only implements DNT by suppressing targeted ads but otherwise continues data collection and use identically, type-1 if data is aggregated, and type-2 if data collection stops all together. If that proves too simplistic, a code could be created from binary values for an ordered set of categories, for example 101 might mean a site collects data for fraud prevention, does not show targeted ads, and does collect data for analytics. It is not necessary to define a standardized code at this early juncture. However, it would be exceedingly helpful to think about what the syntax might look like, and build in a mechanism to support communication.

Eventually, user agents could use this data to inform users about their effective privacy online. However, if we create standards that preclude information flow, we will not be able to build visualization tools later.

I would like to speak at the W3C workshop to highlight the gap between user expectations and what is being built, explain that feedback to users can help communicate across this gap, and encourage discussion about how best to include feedback mechanisms in DNT standards.